From the Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden

RISK PREDICTION MODELS FOR CARDIOVASCULAR DISEASE AND OVERALL MORTALITY

Andrea Ganna

Karolinska Institutet

Stockholm 2015

Published by Karolinska Institutet.
Printed by Printed by Eprint AB 2015

ISBN 978-91-7549-798-3

RISK PREDICTION MODELS FOR CARDIOVASCULAR DISEASE AND OVERALL MORTALITY
THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Andrea Ganna

Principal Supervisor:
Professor Erik Ingelsson
Uppsala University
Department of Medical Sciences
Division of Molecular Epidemiology

Co-supervisor(s):
Professor Marie Reilly
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Doktor Patrik Magnusson
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Opponent:
Professor Paul Eliot
Imperial College London
Department of Epidemiology and Biostatistics

Examination Board:
Doktor Nicola Orsini
Karolinska Institutet
Institute of Environmental Medicine

Professor Johan Hallqvist
Uppsala University
Department of Public Health and Caring Sciences

Professor Lars Alfredsson
Karolinska Institutet
Institute of Environmental Medicine

Abstract

Prediction or prognostication is at the core of modern evidence-based medicine. Prediction of overall mortality and cardiovascular disease can be improved by a systematic evaluation of measurements from large-scale epidemiological studies or by using nested sampling designs to discover new markers from omics technologies.

In study I, we investigated if prediction measures such as calibration, discrimination and reclassification could be calculated within traditional sampling designs and which of these designs were the most efficient. We found that is possible to calculate prediction measures by using a proper weighting system and that a stratified case-cohort design is a reasonable choice both in terms of efficiency and simplicity.

In study II, we investigated the clinical utility of several genetic scores for incident coronary heart disease. We found that genetic information could be of clinical value in improving the allocation of patients to correct risk strata and that the assessment of a genetic risk score among intermediate risk subjects could help to prevent about one coronary heart disease event every 318 people screened.

In study III, we explored the association between circulating metabolites and incident coronary heart disease. We found four new metabolites associated with coronary heart disease independently of established cardiovascular risk factors and with evidence of clinical utility. By using genetic information we determined a potential causal effect on coronary heart disease of one of these novel metabolites.

In study IV, we compared a large number of demographics, health and lifestyle measurements for association with all-cause and cause-specific mortality. By ranking measurements in terms of their predictive abilities we could provide new insights about their relative importance, as well as reveal some unexpected associations. Moreover we developed and validated a prediction score for five-year mortality with good discrimination ability and calibrated it for the entire UK population.

In conclusion, we applied a translational approach spanning from the discovery of novel biomarkers to their evaluation in terms of clinical utility. We combined this effort with methodological improvements aimed to expand prediction measures in settings that were not previously explored. We identified promising novel metabolomics markers for cardiovascular disease and supported the potential clinical utility of a genetic score in primary prevention. Our results might fuel future studies aimed to implement these findings in clinical practice.

List of publications

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals:

I. Ganna A**,** Reilly M, De Faire U, Pedersen NL, Magnusson KE, Ingelsson E. **Risk prediction measures for case-cohort and nested case-control designs: an application to cardiovascular disease.**
American Journal of Epidemiology. 2012; 175(7):715-24

II. Ganna A, Magnusson PK, Pedersen NL, de Faire U, Reilly M, Arnlöv J, Sundström J, Hamsten A, Ingelsson E.
Multilocus Genetic Risk Scores for Coronary Heart Disease Prediction.
Arteriosclerosis, Thrombosis, and Vascular Biology. 2013; 33(9):2267-72

III. Ganna A, Salihovic S, Sundström J, Broeckling CD, Hedman ÅK, Magnusson PKE, Pedersen NL, Larsson A, Siegbahn A, Zilmer M, Prenni J, Ärnlöv J, Lind L, Fall T, Ingelsson E.
Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease.
PLoS Genetics. 2014; 10(12): e1004801

IV. Ganna A**,** Ingelsson E.
Five-year mortality predictors: a prospective study of ~500,000 UK Biobank participants.
Manuscript.

Data, code and results

The results of study IV are also reported at the following website: www.ubble.co.uk.

The code used in study I, II, III and IV is available on the following website: https://github.com/andgan/ Thesis_repository.

Raw data used in study III are available on the following website: http://www.ebi.ac.uk/metabolights/ with accession numbers MTBLS90, MTBLS93 and MTBLS124.

Related publications
(not included in the thesis)

- Ganna A, Rivadeneira F, Hofman A, Uitterlinden AG, Magnusson PK, Pedersen NL, Ingelsson E, Tiemeier H.
  Genetic determinants of mortality. Can findings from genome-wide association studies explain variation in human mortality?
  Human Genetics. 2013; 132(5):553-61

- Berndt SI*, Gustafsson S*, Mägi R*, Ganna A*, Wheeler E, Feitosa MF, Justice AE, Monda KL, Croteau-Chonka DC, Day FR, Esko T, Fall T, Ferreira T, Gentilini D, Jackson AU, Luan J, Randall JC, Vedantam S, Willer CJ, Winkler TW, Wood AR, Workalemahu T, Hu YJ, Lee SH, Liang L, Lin DY, Min JL, Neale BM, Thorleifsson G, Yang J, Albrecht E, Amin N, Bragg-Gresham JL, Cadby G, den Heijer M, Eklund N, Fischer K, Goel A, Hottenga JJ, Huffman JE, Jarick I, Johansson Å, Johnson T, Kanoni S, Kleber ME, König IR, Kristiansson K, Kutalik Z, Lamina C, Lecoeur C, Li G, Mangino M, McArdle WL, Medina-Gomez C, Müller-Nurasyid M, Ngwa JS, Nolte IM, Paternoster L, Pechlivanis S, Perola M, Peters MJ, Preuss M, Rose LM, Shi J, Shungin D, Smith AV, Strawbridge RJ, Surakka I, Teumer A, Trip MD, Tyrer J, Van Vliet-Ostaptchouk JV, Vandenput L, Waite LL, Zhao JH, Absher D, Asselbergs FW, Atalay M, Attwood AP, Balmforth AJ, Basart H, Beilby J, Bonnycastle LL, Brambilla P, Bruinenberg M, Campbell H, Chasman DI, Chines PS, Collins FS, Connell JM, Cookson WO, de Faire U, de Vegt F, Dei M, Dimitriou M, Edkins S, Estrada K, Evans DM, Farrall M, Ferrario MM, Ferrières J, Franke L, Frau F, Gejman PV, Grallert H, Grönberg H, Gudnason V, Hall AS, Hall P, Hartikainen AL, Hayward C, Heard-Costa NL, Heath AC, Hebebrand J, Homuth G, Hu FB, Hunt SE, Hyppönen E, Iribarren C, Jacobs KB, Jansson JO, Jula A, Kähönen M, Kathiresan S, Kee F, Khaw KT, Kivimäki M, Koenig W, Kraja AT, Kumari M, Kuulasmaa K, Kuusisto J, Laitinen JH, Lakka TA, Langenberg C, Launer LJ, Lind L, Lindström J, Liu J, Liuzzi A, Lokki ML, Lorentzon M, Madden PA, Magnusson PK, Manunta P, Marek D, März W, Mateo Leach I, McKnight B, Medland SE, Mihailov E, Milani L, Montgomery GW, Mooser V, Mühleisen TW, Munroe PB, Musk AW, Narisu N, Navis G, Nicholson G, Nohr EA, Ong KK, Oostra BA, Palmer CN, Palotie A, Peden JF, Pedersen N, Peters A, Polasek O, Pouta A, Pramstaller PP, Prokopenko I, Pütter C, Radhakrishnan A, Raitakari O, Rendon A, Rivadeneira F, Rudan I, Saaristo TE, Sambrook JG, Sanders AR, Sanna S, Saramies J, Schipf S, Schreiber S, Schunkert H, Shin SY, Signorini S, Sinisalo J, Skrobek B, Soranzo N, Stančáková A, Stark K, Stephens JC, Stirrups K, Stolk RP, Stumvoll M, Swift AJ, Theodoraki EV, Thorand B, Tregouet DA, Tremoli E, Van der Klauw MM, van Meurs JB, Vermeulen SH, Viikari J, Virtamo J, Vitart V, Waeber G, Wang Z, Widén E, Wild SH, Willemsen G, Winkelmann BR, Witteman JC, Wolffenbuttel BH, Wong A, Wright AF, Zillikens MC, Amouyel P, Boehm BO, Boerwinkle E, Boomsma DI, Caulfield MJ, Chanock SJ, Cupples LA, Cusi D, Dedoussis GV, Erdmann J, Eriksson JG, Franks PW, Froguel P, Gieger C, Gyllensten U, Hamsten A, Harris TB, Hengstenberg C, Hicks AA, Hingorani A, Hinney A, Hofman A, Hovingh KG, Hveem K, Illig T, Jarvelin MR, Jöckel KH, Keinanen-Kiukaanniemi SM, Kiemeney LA, Kuh D, Laakso M, Lehtimäki T, Levinson DF, Martin NG, Metspalu A, Morris AD, Nieminen MS, Njølstad I, Ohlsson C, Oldehinkel AJ, Ouwehand WH, Palmer LJ, Penninx B, Power C,

Province MA, Psaty BM, Qi L, Rauramaa R, Ridker PM, Ripatti S, Salomaa V, Samani NJ, Snieder H, Sørensen TI, Spector TD, Stefansson K, Tönjes A, Tuomilehto J, Uitterlinden AG, Uusitupa M, van der Harst P, Vollenweider P, Wallaschofski H, Wareham NJ, Watkins H, Wichmann HE, Wilson JF, Abecasis GR, Assimes TL, Barroso I, Boehnke M, Borecki IB, Deloukas P, Fox CS, Frayling T, Groop LC, Haritunian T, Heid IM, Hunter D, Kaplan RC, Karpe F, Moffatt MF, Mohlke KL, O'Connell JR, Pawitan Y, Schadt EE, Schlessinger D, Steinthorsdottir V, Strachan DP, Thorsteinsdottir U, van Duijn CM, Visscher PM, Di Blasio AM, Hirschhorn JN, Lindgren CM, Morris AP, Meyre D, Scherag A, McCarthy MI*, Speliotes EK*, North KE*, Loos RJ*, Ingelsson E*.
Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture.
Nature Genetics. 2013; 45(5):501-12

- Ganna A, Ortega-Alonso A, Havulinna A, Salomaa V, Kaprio J, Pedersen NL, Sullivan PF, Ingelsson E, Hultman CM, Magnusson PKE.
Utilizing twins as controls for non-twin case-materials in genome wide association studies.
PLoS One. 2013; 8(12):e83101

- Ganna A*, Lee D*, Ingelsson E, Pawitan Y.
Rediscovery rate estimation for assessing the validation of significant findings in high-throughput studies.
Briefings in Bioinformatics. 2014*;*

- Ganna A*, Fall T*, Lee W, Broeckling CD, Kumar J, Hägg S, Magnusson PKE, Prenni J, Lind L, Pawitan Y and Ingelsson E.
A workflow for UPLC-MS non-targeted metabolomic profiling in large human population-based studies.
Preprint available at *http://biorxiv.org/content/early/2014/02/19/002782*

* Indicates equal contribution

Contents

List of abbreviations

AUC	Area under the receiver operating characteristic curve
BMI	Body mass index
CHD	Coronary heart disease
CVD	Cardiovascular disease
DNA	Deoxyribonucleic acid
ECG	Electrocardiography
EDTA	Ethylenediaminetetraacetic acid
FHS	Framingham Heart Study
GWAS	Genome wide association studies
IPT	In-person testing
MGRS	Multilocus genetic risk score
MS	Mass spectroscopy
NMR	Nuclear magnetic resonance
NRI	Net reclassification improvement
ROC	Receiver operating characteristic
SNP	Single nucleotide polymorphism

1 Introduction

1.1 The concept of prediction or prognostication

The concept of prediction or prognostication is central to modern evidence-based medicine. Clinical prediction models are currently used in primary or secondary prevention to predict diseases or unfavourable outcomes or to determine the best choice of therapy. Evidence-based medicine will undoubtedly see an increase of these tools in the future, together with advancements in the different fields required for their development.

Nevertheless, prediction has always been practiced in medicine as much as prediction abilities have always been in the skillset of clinicians. However, the quantification and rationalization of what has often been called ‘intuition’ was possible only with the advent of data-rich clinical studies and appropriate statistical methodologies.

Statistics is the backbone of prediction models. Statistical models are suitable for combining results obtained from different technologies, such as genetics, proteomics or metabolomics into a single predictive equation that returns meaningful information for clinicians and patients.

Importantly, having well-defined statistical models allows rational and shared decision-making. Further, such quantifiable decisions can be evaluated at public policy levels for planning specific intervention strategies. Finally, information derived from prediction models is important to make individuals aware of their conditions in order to prevent future diseases.

Following the pioneering work conducted on participants from the Framigham Heart study (FHS) [1], cardiovascular disease research has become the preferred field for developing prediction models and studying their utility for decision-making purposes.

2 Cardiovascular disease

2.1 Definition of cardiovascular disease

The term cardiovascular disease includes diseases of the heart, brain and peripheral vessels. Pathological processes affecting the heart may be broadly divided into three major categories:

1. Disease of the cardiac muscle commonly defined as cardiomyopathies.
2. Disease of the conduction system that lead to disturbance of the heart rhythm, called arrhythmias.
3. Disease affecting the coronary arteries, called atherosclerosis.

This thesis mainly focuses on the third category, atherosclerotic cardiovascular disease, which also includes ischemic stroke, and from now on is referred as CVD.

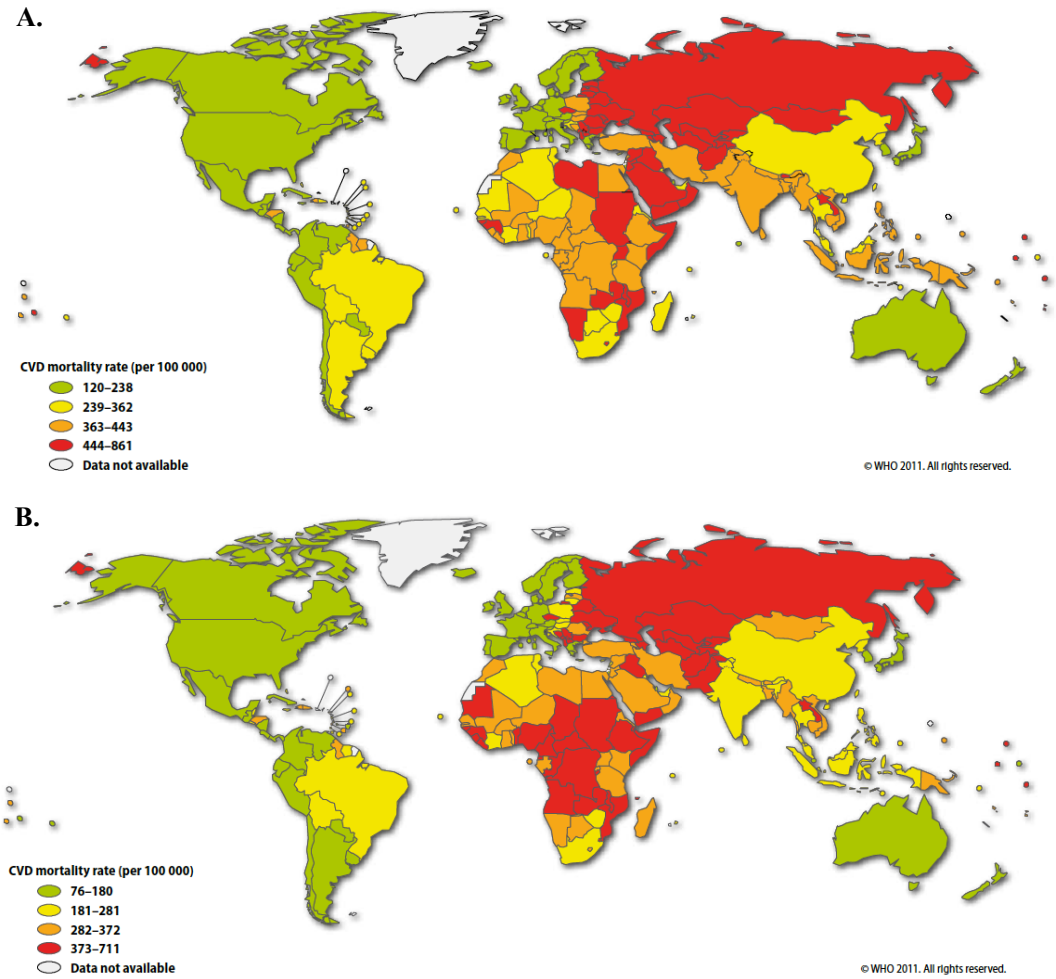
2.2 Epidemiology of cardiovascular disease

CVD, including coronary heart disease (CHD) and stroke remains the leading cause of mortality worldwide, accounting for approximately 30% of total mortality [2]. Moreover, this unfortunate leadership is expected to last at least to the year 2030 [3].

Although there has been a decline in CVD mortality in high-income countries over the years, the burden of this disease is increasing as a result of adaptation of a Westernized lifestyle in poorer countries with increasing obesity, type 2 diabetes and smoking rates (see **Figure 1**) [4] .

Finally, the non-fatal CVD events contribute significantly to overall disability, estimated by the disability-adjusted life years. In 2010, the total burden of CVD was 295 millions disability-adjusted life years, reaching an increase of 22.6% compared to 1990 [5].

Figure 1. *World map showing the global distribution of CVD mortality rates (age standardized, per 100,000) in males (panel A) and females (panel B)* [6].



2.3 Etiology of cardiovascular disease

CVD is a multifactorial disease, with a number of modifiable physiological risk factors such as high blood pressure [7], high total cholesterol [8], high blood glucose [9] and high body mass index (BMI) [10].

Also modifiable behavioural risk factor play a causal role and include increased alcohol use [11], tobacco smoking [12], unhealthy diet [12] and physical inactivity [13]. Finally, familial aggregation of CVD suggests evidence of a genetic predisposition [14] and twin studies have reported about 40% heritability of CHD mortality [15].

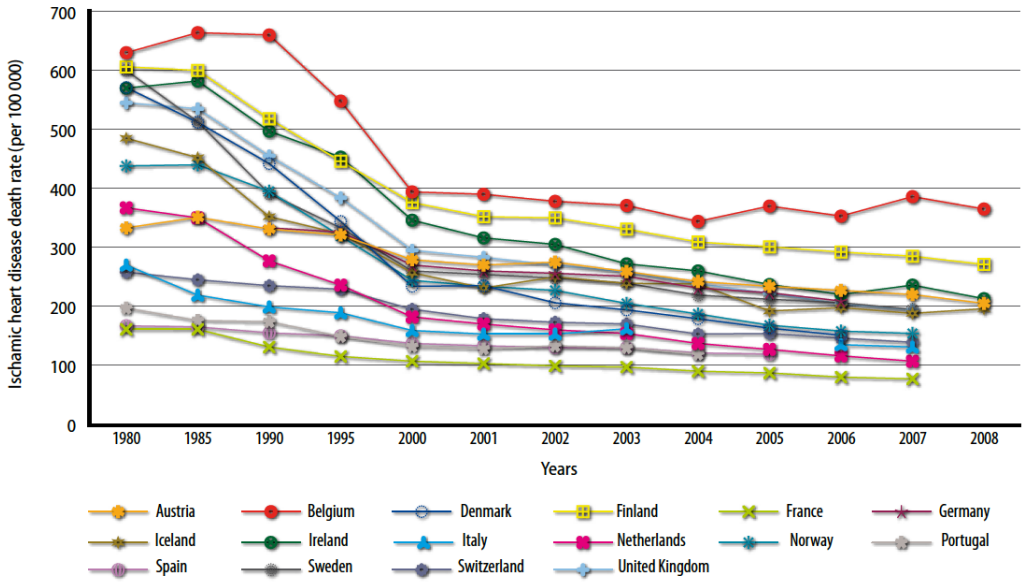
2.4 Challenges in cardiovascular disease prevention

Most of the cardiovascular risk factors where already established before 1975 [16] and, in the following decades, improvement in primary prevention, cardiac care and major pharmaceutical drugs have contributed to dramatically decrease CVD mortality (see **Figure 2**) [17].

Nevertheless, this decrease has not been homogenous between all CVD subtypes. For example, an improvement in survival has been observed for myocardial infarction, but not for stroke or heart failure. Moreover, although efforts to decrease cholesterol levels and smoking have been successful in the United States and Europe over the past two decades, prevalence of obesity and type 2 diabetes is on the rise and only ~1% of individuals in the United States reach what the American Heart Association has defined "ideal cardiovascular health" [18, 19].

Is therefore important, on one side, to identify new strategies (pharmacological or behavioural) for reducing the CVD burden in the community and, on the other, to identify the right targets for such interventions. Prediction models can be used to identify and assign individuals to the best intervention strategy and are therefore a key component of the challenge of reducing the CVD burden.

Figure 2. *Trends in CVD mortality rates (age standardized) in developed countries* [6].



3 Prediction models

3.1 Study design for building prediction models

Risk prediction and prognosis are inherently longitudinal in nature. Therefore, prospective cohort studies with participants followed over time are a natural choice for evaluating these aspects. Case-control studies are not suitable to evaluate and develop prediction models, while case-cohort and nested case control designs can be used, but attention to several methodological aspects needs to be paid [20].

An optimal study should have a high participation rate and should be conducted in a sample of the population for which the prediction model is intended. In addition, the study should also be able to follow up patients for both mortality and incidence of the disease of interest. Moreover, if the aim is to build a new prediction model, the study should preferably collect a large number of reproducible measurements, not necessary causally related with the disease.

3.2 Model performances: Discrimination

Discrimination relates to how well a prediction model can discriminate those with the outcome from those without the outcome. A model that places each individual in the class to which he or she truly belongs would be said to have a perfect discrimination.

One of the most popular measures of discrimination used in the context of dichotomous outcomes is derived from the receiver operating characteristic (ROC) curve. The ROC curve is the entire set of possible true and false positive fractions (i.e. sensitivity and 1- specificity) attainable by dichotomizing the outcome of interest with different thresholds [21].

The most widely used summary measure is the area under the ROC curve (AUC), also known as the C-statistic. A perfect test, which always assigned individuals to the correct outcome, has a AUC value of one. Conversely, an uninformative test, has AUC=0.5. The AUC can be interpreted as the probability that a randomly selected person with the event has a higher predicted risk than a randomly selected person without the event.

Generalized versions of the AUC for survival analysis have also been developed [22, 23]. The most commonly used measure is the Harrell's C-index [24], which measures the concordance between the predicted survival time and the actual survival time in

pairs of individuals where at least one has experienced the event of interest.

Finally, differences in AUC or C-index between two models that share all risk factors, except for a new marker, can be used to evaluate the improvement in discrimination due to the new marker. However, in this setting the AUC has been criticized to be insensitive to clinically important risk differences [25]. Moreover, a formal test for AUC improvement [26] has been shown to be underpowered compared to more common tests to compare nested regression models (e.g. Wald or likelihood-ratio test) [27].

3.3 Model performances: Calibration

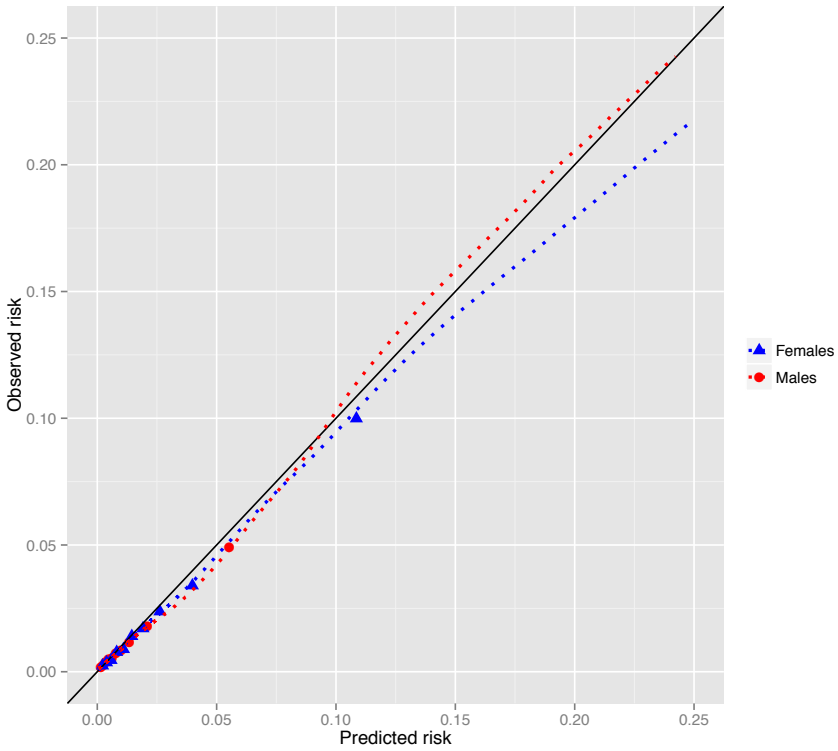
Calibration measures how closely the predicted risk approaches the observed risk. More specifically, it is the agreement between the probability of developing the outcome of interest within a certain time-period as estimated by the prediction model and the observed outcome frequencies.

To assess the model's calibration, a plot of the predicted and observed events (or probabilities of events) can be used. Perfect prediction should follow the $45°$ line. A deviation from this line indicates overestimation or underestimation of the true number of events. In **Figure 3**, we report an example of calibration plot obtained from study IV.

This aspect can be formally tested by comparing observed and expected number of events in each risk categories as defined by risk deciles. Such χ^2 test is called Hosmer-Lemeshow test [28]. A similar test for survival data has also been proposed (Grønnesby and Borgan goodness-of-fit test) [29]. However, these tests do not capture all the potential miscalibration patterns. Moreover, a test might suggest good calibration simply due to lack of power.

Good calibration does not mean good discrimination and vice versa. A model, once tested in a different population from the one where it has been developed, can still have good discrimination, but the calibration might be poor because the disease incidence is different. A processed known as re-calibration can be used to re-fit the model to the new population [30].

Figure 3. *Example of calibration plot of observed vs. predicted risk obtained from study IV. The deciles of risk are represented by the triangles and the dotted lines are the loess smoothers. Overlap with the 45° line indicates perfect calibration.*



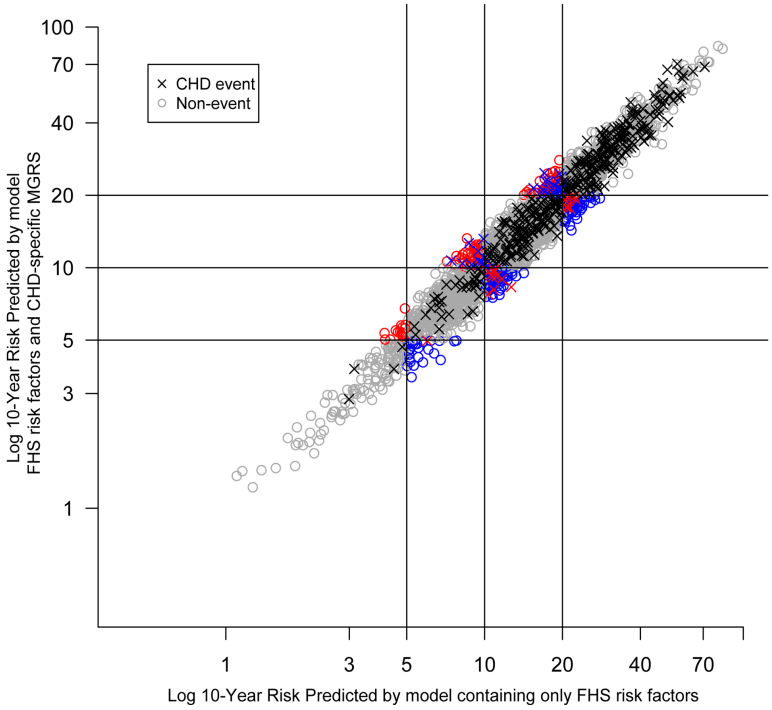
3.4 Reclassification

An important property of a prediction model is the ability to stratify the population into clinically relevant risk categories. A perfect model would assign the entire population to the very highest or very lowest risk categories and leave no one in the middle categories, in which there are still uncertainties about the appropriate course of action.

Reclassification measures are typically used to evaluate this property when a new risk marker is added to a risk prediction model that includes an established set of markers. The most commonly used measure of reclassification, net reclassification improvement (NRI), was introduced by Pencina and colleagues [31]. This approach focuses on reclassification tables constructed separately for individuals with and without the event of interest and quantifies the correct movement in categories: upwards for those that eventually experience an event and downwards for those that do not experience an event. In **Figure 4**, we report an example of reclassification plot and table from study II.

NRI is meaningful when risk categories are well defined (e.g. by guidelines) and when specific behavioural or pharmacological interventions are available for high-risk patients. The NRI has been criticized because it is highly dependent on the number of risk categories chosen [32], lacks clinical interpretation [33] and is difficult to interpret in some situations [34].

Figure 4. *Reclassification graph and table obtained from study II representing the individual risk at 10 years for a model with FHS risk factors vs. individual risk at 10 years for a model with FHS risk factors + multilocus genetic score (MGRS) for CHD. Blue marks are correctly reclassified, red marks are incorrectly reclassified between categories of clinical interest (<10%, 10% to 20%, >20%).*



Original risk categories	**Reclassification**			
	≤5%	6% to ≤10%	11% to ≤20%	>20%
	number of participants			
Participants without CHD events				
≤5%	116	17	0	0
6% to ≤10%	26	593	84	0
11% to ≤20%	0	109	946	63
>20%	0	0	92	580
Participants with CHD events				
≤5%	4	0	0	0
6% to ≤10%	1	31	10	0
11% to ≤20%	0	11	123	20
>20%	0	0	8	180

Net reclassification for participants without CHD: 63 of 2626 (2.4%).
Net reclassification for participants with CHD: 10 of 388 (2.6%) [a].

3.5 Measures of clinical utility

In addition to evaluate the performances of the prediction model using measures such as discrimination and calibration, it is also important to determine if the model is clinically useful. The NRI partially tackles this problem by introducing clinically meaningful risk categories in the measure assessment. However, it does not touch upon the concept of benefit and harm, which are at the core of medical decision-making [35].

Vickers and Elkin proposed a decision-curve analysis as a simple approach to quantify the clinical usefulness of a prediction model (or an extension to a model) [36]. This method is used to calculate the net benefit, which is the weighted difference between benefits and harms (defined as true positive and false positive decisions, respectively).

Rapsomaniki and colleagues introduced a different version of the net benefit, which quantifies the number of event-free life years saved per 1,000 people screened. Differently from the decision-curve analysis approach, external information, including the cost of treatment per person per year and the monetary value of event-free life years, needs to be included [37].

Finally, prediction models can be evaluated using health-economic approaches, such as Markov chain Monte-Carlo methods [38]. These models can be used to perform cohort simulations as well as individual-based microsimulations and require numerous external information, which is normally obtained from the literature or from existing population-based studies.

3.6 Overfitting and validation

Performance measures that are calculated in the same data sample from which the prediction model is developed tend to perform overoptimistically.

This is simply because the model is designed to fit the data, but becomes less accurate when tested in new but similar individuals. This concept is called overfitting. The overfitting is greater with a smaller sample size and when a large number of measurements are included in the model.

There are several strategies to control the overfitting and they are traditionally divided in internal and external validation.

Internal validation aims to reduce the overfitting by using the same population that generated the model. Split-sample and *k*-fold cross-validation are commonly used

strategies to divide the study in specific subsamples, some of which are used to develop the model and the others to validate it. In the case of *k*-fold cross-validation this process is repeated iteratively.

External validation reduces the overfitting by using individuals that differ in some aspects from those in which the model is developed. External validation can be applied to individuals collected at a different time points (temporal validation), in individuals from a different geographical regions (geographical validation) or in a completely separate population, collected by independent investigators.

3.7 Prediction models in cardiovascular disease

Starting with the Framingham risk score over 30 years ago [39], the cardiovascular disease field has been a fruitful source of prediction models.

There are several reasons behind this favorable attitude:

1. Cardiovascular disease is the leading cause of death and therefore a common outcome in the population. This makes feasible to plan prospective population-based studies and to observe enough events within a limited follow-up time (often 10 years).
2. Risk factors for cardiovascular disease are well known and easily measurable. Those can be combined in a prediction model with relatively high prediction performances.
3. Individuals that are likely to develop a cardiovascular event, as detected by prediction models, can be assigned to several preventive strategies, both pharmacological and behavioural, without major side effects.

The original Framingham risk score was developed in 5,209 men and women between the ages of 30 and 62 recruited in 1948 from the town of Framingham, Massachusetts. The primary outcome was prediction of eight-year CVD and included the following risk factors: age, systolic blood pressure, total cholesterol, glucose intolerance, cigarette smoking and left ventricular hypertrophy by electrocardiogram [39]. In **Figure 5**, we report the original coefficients published in the 1976 paper, including the description on how to calculate the score.

Figure 5. *Original table from Kannel's 1976 paper describing a first version of the Framingham risk score* [39].

TABLE IV

Coefficients for Calculating Risk of Cardiovascular Disease

Variable	Coefficient	
	Men	Women
Age (years)	0.3743307	0.2665693
Age × age	−0.0021165	−0.0012655
Serum cholesterol (mg/ml)	0.0258102	0.0160593
Systolic blood pressure (mm Hg)	0.0156953	0.0144265
Cigarette smoking*	0.5583013	0.0395348
LVH by ECG*	1.0529656	0.8745090
Glucose intolerance*	0.6020336	0.6821258
Cholesterol × age	−0.0003619	−0.0002157
Intercept	−19.7709560	−16.4598427

* Yes = 1, no = 0 (for definitions see Shurtleff et al.[2]).

To obtain the probability that cardiovascular disease will occur in 8 years to a man or woman initially free of cardiovascular disease multiply the value of the characteristic in the units specified by the coefficient for the variable, sum these products and add the intercept. This provides the coefficient (C) to calculate the probability, $P = 1/(1 + e^{-C})$.

ECG = electrocardiogram; LVH = left ventricular hypertrophy.

Different risk factors have been included in following versions of the score [40, 41] and extensions to evaluate the risk of ten-year atrial fibrillation [42], ten-year CHD [43], 30-year CVD [44], eight-year diabetes [45], four-year hypertension [46] and ten-year stroke [47] have also been proposed.

Several cardiovascular prediction models have been developed in Europe because the Framingham risk scores were not entirely suitable to the general European population. The SCORE equation has been recommended by the Fifth Joint European Task Force on cardiovascular prevention [48]. The QRISK2 algorithm was developed using a population-based clinical research database in the UK [49]. A simple PROCAM score scheme [50] and a neural network model [51] were developed in Germany. An ASSIGN risk score involving family history of CVD was developed in Scotland based on the Scottish Heart Health Extended Cohort [52] and finally, a CUORE equation was developed in Italy for a low coronary incidence population [53].

Similarly, other CVD risk models have been constructed elsewhere in the world. A MUCA ischemic CVD risk model was developed for the Chinese population [54]. A multivariate regression model involving C-reactive protein was developed in Japan [55]. A recalibrated Framingham model was investigated in Thailand [56] and a

multivariate risk prediction model for CHD was developed in Australia based on the Busselton Health Study [57].

3.8 Prediction models for overall mortality

Similarly to what done for CVD, prediction models have also been developed for overall mortality. Although mortality is a very heterogeneous trait, identification of high-risk individuals with reduced life expectancy is important from a public health perspective. Guidelines are increasingly incorporating life expectancy as a central factor in weighing the benefits and the burdens of tests and treatments.

For example, the US Preventive Services Task Force recommends routine colorectal cancer screening for adults aged 50 to 75 years [58]. One reason for using a 75 year old cut-off is that the average life expectancy for 75-year-old US adults (11.1 years in 2000) is similar to the time to benefit for colorectal cancer screening (10.3 years) [59]. In this case, the life expectancy is simply obtained on a population level (e.g. from life tables). However, it is clear that life expectancy is heterogeneous among 75 years old individuals, depending on their individual risk profile, and the calculation of life expectancy simply based on age is an oversimplification. Therefore, the use of prognostic indices in this setting might improve the decision-making about preventive strategies [60]. However, it has also been highlighted that the accuracy and generalizability of current prediction tools is low [61].

There are several reasons why building an appropriate prediction model of overall mortality might be challenging:

1. Mortality is not uniformly distributed across ages. Among younger individuals, death due to diseases is a relatively rare event and therefore large population studies are needed to gather enough events in a short follow-up time.
2. Risk factors for overall mortality are heterogeneous and, unlikely those for CVD, there is not a consensus regarding which factors to include in a prediction model. Comorbidities are important predictors of mortality, but prediction models only based on previous diseases or conditions might encounter problems due to reporting and diagnostic biases. Lifestyle measurements and biomarkers are preferred risk factors, but they are often difficult to collect on large-scale.
3. The strongest predictors of mortality are often geographical and time-dependent. This makes the generalizability of the results challenging, especially if the study is not contemporary or representative of the general population.

Nevertheless several prediction models for mortality have been developed. The Charlson index is the one of the most popular comorbidities-based tools to predict 10-year mortality [62].

Several prediction models have been developed in older adults. A recent review by Yourman and colleagues summarized these models in terms of potential for bias, generalizability, and accuracy [61]. Overall, the authors identified 16 validated prediction models. Eight were developed in hospital patients, six in community-dwelling patients and two in nursing home patients. Prediction models were mostly developed in participants from US, while only four from Europe. Only two models were validated by investigators not involved in the studies' development, and no model was prospectively tested and found to be accurate in a large diverse sample. They concluded that no study was completely free from potential sources of bias.

4 Omics and prediction

4.1 Genomics

Genomics has known a phase of intense development in the past years. The first sequence of the human genome in 2001 [63, 64] and the subsequent HapMap Project [65] have helped to shed light into the characteristics and the haplotype structure of the human genome and led to the explosion of genome-wide association studies (GWAS). The assumption behind these studies is the so-called 'common disease–common variant' hypothesis, which posited that common genetic variants could have a role in the etiology of common diseases [66]. Thus, by an unbiased scan of common variants in the genome one could, in principle, pinpoint key genes and help to outline underlying mechanisms.

Specifically, GWASs take advantage of the principle of linkage disequilibrium at population level to identify genetic markers to tag a haplotype, and the number of such tagging markers is much smaller than the total number of segregating variants in the population. With this strategy the cost-per-sample can be reduced and a large number of individuals can be genotyped for these tagging markers, also called single nucleotide polymorphisms (SNPs). For example, a selection of approximately 500,000 common SNPs in the human genome is sufficient to tag common variation.

Simple statistical models combined with a tight control for the number of false-positives (often using a P-value $< 10^{-8}$ to declare significance) and replication in external populations can be used to compare cases of a specific disease and control in several thousands individuals across hundred thousands to millions SNPs. The large number of individuals is needed in order to detect the small effect of SNPs on the disease with sufficient statistical confidence.

In the recent years, hundreds of these studies has been conducted and have identified SNPs associated with more than 1,000 traits or diseases. If, on one side, these discoveries can be used to understand the biological basis of diseases and to potentially develop new pharmacologic therapies, on the other, results from GWASs have the potential to be useful for disease prediction.

However, given the small effect sizes observed for most of these associations and the relative small portion of the heritability explained, it has been argued that the clinical impact of genetic risk prediction using common SNPs is limited [67-70]. Moreover,

several pitfalls in making predictions of common diseases or complex traits from genetics data have been identified. These include errors in estimating the effect size of the markers, inappropriate statistical methods, lack of validation, overfitting and population stratification [71].

4.2 Metabolomics

Metabolomic profiling, or metabolomics, can be described as a holistic approach to the study of low-weight molecules (<1,500 Daltons) called metabolites. These molecules are produced by chemical processes in the body (i.e. metabolism) or from exogenous sources (e.g. diet, drugs, xenobiotics or gut-host co-metabolism). Metabolites are often measured in blood or urine, but they can also be detected in saliva, breath or any of the approximately 500 different histological cell types in the human body [72].

Improvements in instrumental technologies and advances in bioinformatics tools have provided the possibility to perform metabolomics on large prospective epidemiological studies with thousands of individuals and hundreds of phenotypes [73].

The two main approaches to perform large-scale metabolomics studies are nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry (MS). NMR has high reproducibility, requires little sample preparation and the cost per sample is relatively low. However, it is less sensitive, meaning that fewer metabolites can be detected. Mass spectrometry allows the detection of a larger number of metabolites, but it is less reproducible, more platform dependent and the analysis of the generated data is more complex.

NMR and MS can run either in targeted or non-targeted mode. The targeted approach relies on the measurement of a specific subset of metabolites, typically focusing on pathways of interest. However, with this approach a large number of disease-related metabolites are likely to be missed. The non-targeted approach has the advantage to simultaneously measure as many metabolites as possible from a biological sample. This approach however requires a post-hoc annotation of the results and, often, the chemical structure of some metabolites cannot be resolved.

Disease prediction using metabolomics is a relatively new field and few promising studies have been conducted to investigate this aspect in relation to cardiovascular disease [74-76].

4.3 Challenges in measuring omics technology on large scale: The role of sampling designs

Recent technological developments have allowed researchers to assess thousands of genetic markers, proteins, and metabolites in small amounts of biologic specimens. In parallel, several new large initiatives in different countries have started to collect baseline information and biological specimens from hundreds of thousands of individuals [77, 78]. The combination of huge study samples and high costs for these new technologies makes it unfeasible to measure these new markers on an entire study population, so there is a clear need for efficient study designs.

Researchers are often interested in selecting from an on-going prospective cohort study all the cases of a disease and an appropriate set of controls. Thus, the markers of interest are measured only in this subsample of participants. Commonly used sampling designs are:

1. Case cohort. All cases of a specific disease or of multiple diseases are collected and controls are sampled from original cohort (subcohort) independently of the definition of the cases.
2. Nested case-control. All cases for one specific disease are collected and controls are sampled from individuals at risk at the times when cases are identified.

These sampling designs have been used in several studies [79-82] to investigate the prediction abilities of new cardiovascular markers above established risk factors. Nevertheless, methodological investigations regarding the appropriate way to derive measures of prediction performances within a sampling design setting are lacking.

5 Aims

The general aim of this thesis is to evaluate established measurements and novel markers derived from omics technologies in terms of their ability to predict cardiovascular disease and overall mortality, as well as to investigate methodological aspects related to the use of prediction metrics in epidemiological studies.

The specific aims of this thesis, each corresponding to one of its four component papers, are:

- To investigate the ability to adequately estimate individual risk and risk prediction metrics in unstratified and stratified (matched) case-cohort and nested case-control designs.
- To investigate the clinical utility of several genetic scores for CHD.
- To discover new markers of CHD by analysing circulating metabolites and to determine if these metabolites can improve CHD prediction beyond established risk factors.
- To compare a large number of potential predictors of five-year mortality and to develop and validate a prediction score for five-year mortality using only self-reported information.

6 Study populations

6.1 Swedish Twin Registry: Overview

The Swedish Twin Registry is a nation-wide register including over 194,000 Swedish twins born from 1886 to 2008 [83]. Since 2004, all identified nine-year-old have been contacted and invited to participate to an ongoing study [84]. Before 2004, the twins were contacted at various ages depending on birth-year cohorts. Over the past ten-years, DNA information was obtained from 45,000 twins, 15,000 of which also provided blood samples.

6.2 Swedish Twin Registry: TwinGene

TwinGene is a longitudinal study nested within the Swedish Twin Register initiated to examine associations between genetic factors and cardiovascular disease in Swedish twins. Twins born before 1958 and who has participated in a telephone screening between 1998 and 2002 were re-contacted between April 2004 and December 2008. Health and medication data were collected from self-reported questionnaires, and a blood sampling kit was mailed to the subject who then contacted a local health care center for blood sampling and a health check-up. Contacts were allowed on Monday to Thursday mornings (not the day before a national holiday), to ensure that the sample would reach the KI Biobank in Stockholm the following day by over night mail. The participants were instructed to fast from 8 PM (20:00) the previous night. A total volume of 50 ml of blood was drawn from each individual by venipuncture.

First, a tube containing Ethylenediaminetetraacetic acid (EDTA) was filled and inverted five times immediately. These plasma samples were used for DNA extraction. Second, three gel tubes were filled, inverted five times immediately, let stand for 30 minutes for coagulation in room temp, centrifuge for 10-15 minutes at 3800 rpm. Thus, the serum was tapped from the gel tubes to a collection tube and placed in a transport cylinder. These serum samples were used for storage and for the following metabolomics analysis. Serum samples for biochemical analysis were collected in the same way.

Tubes were sent to Karolinska University Laboratory by overnight post where they were frozen at -80° C in liquid nitrogen until analysis.

In total, 12,591 individuals (55% women) participated by donating blood to the study, and by answering questionnaires about lifestyle and health.

In study I, we included 6,558 unrelated individuals (only one twin per pair) with cardiovascular risk factors measurements available.

In study II, we included in the analyses all genotyped twins with cardiovascular risk factors measurements available and without previous CHD events, resulting in 7,597 individuals.

Metabolomics was performed in a subsample of TwinGene. Specifically, we utilized a case-cohort design by selecting all the incident cases of coronary heart disease, type 2 diabetes, ischemic strokes and dementia up to 31st December 2010 and a sub-cohort (controls) of 1,643 individuals (43% women). The subcohort was stratified on median age and sex, and for each of the four strata, we randomly selected a number of participants proportional to the corresponding number of cases. Thus, in study III we included the participants sampled in the sub-cohort and the incident CHD cases for a total of 1,670 individuals with cardiovascular risk factors measurements available and without previous CHD events at baseline.

6.3 Swedish Twin Registry: SATSA

The SATSA sample comprises all pairs of twins who indicated that they had been separated before the age of 11 and reared apart, and a sample of twins reared together matched on gender, date and county of birth [85]. SATSA twins aged 50 and older were invited to participate in in-person testing (IPT) sessions in which questionnaires including items concerning health-related behaviours (e.g. alcohol, tobacco, and dietary habits), cognitive tests and physical health measures were administrated. SATSA twins have been followed longitudinally with up to nine IPT sessions across 24 years. In study II, samples for DNA extraction, lipid measurements and other cardiovascular risk factors from the third IPT (1992-1994) were used. The SATSA study was used in study II, in total, 435 individuals with cardiovascular risk factors measurements available and without previous CHD events were included in the analyses.

6.4 Swedish Twin Registry: OCTO-Twin

The OCTO-Twin sample included all twin pairs in Sweden aged 80 years or older in 1991-1994 (i.e. birth years 1913 or earlier) [86]. Up to five waves of IPT sessions at two-year intervals were conducted on all living twins who agreed to participate,

irrespective of co-twin's vital status. Blood samples for subsequent extraction of DNA were collected between the first and second IPT. Questionnaires were similar to the ones used in SATSA. Cardiovascular risk factors were measured at the first IPT (1991-1994). The OCTO-Twin study was used in study II of this thesis, in total, 410 individuals with cardiovascular risk factors measurements available and without previous CHD events were included in the analyses.

6.5 Swedish Twin Registry: GENDER

The GENDER sample included all living pairs of unlike-sex twins born between 1906 and 1925 [87]. Surveys assessing health and other factors were sent to the twin pairs. A subset of this population-based sample aged 70-79 years completed IPT similar to those in SATSA and OCTO-Twin. Three waves of IPT sessions were carried out at four years intervals during 1995-2004. Blood samples for DNA extraction and cardiovascular risk factors were collected during the first IPT session (1995-1997). The GENDER study was used in study II of this thesis. In total, 421 individuals with cardiovascular risk factors measurements available and without previous CHD events were included in the analyses.

6.6 Swedish Twin Registry: HARMONY

All twins from the Swedish Twin Registry aged 65 and older were screened by telephone for cognitive dysfunction [88]. This included any surviving twins from the SATSA, OCTO-Twin, and GENDER studies described above. Among those screened, 11.5% were positive for suspicion of dementia and were referred for complete clinical evaluation and blood sampling by a physician or a nurse (1999-2001). Once the preliminary IPT suggested dementia, the twin partner was also invited for an identical clinical work-up. In study II, 936 participants not recruited in SATSA, OCTO or GENDER with genotype information available were considered. The HARMONY study was used in study II. In total, 767 individuals with cardiovascular risk factors measurements available and without previous CHD events were included in the analyses.

6.7 ULSAM

Men born between 1920 and 1924 in Uppsala, Sweden were invited to participate at age 50 (N=2,841) in this longitudinal cohort study, which was started in 1970 [89]; 81.7% (N=2,322) participated. Individuals were re-investigated at the ages of 60, 70,

77, 82 and 88 years. Information collected includes a medical questionnaire, blood pressure and anthropometric measurements, glucose tolerance test and 24-hour ambulatory blood pressure.

At age 70, EDTA plasma, citrate plasma, serum and whole blood for DNA extraction were collected from fasting participants and stored at -70° C in liquid nitrogen until analysis. Additional EDTA plasma was collected during oral glucose tolerance test. EDTA plasma samples were used for the metabolomics profiling.

In study II, we included in the analyses, all individuals genotyped at age 70 with cardiovascular risk factors measurements available and without previous CHD events, resulting in 981 individuals.

In study III, 1,028 individuals with cardiovascular risk factors and metabolomics measurements available and without previous CHD events were included in the analyses.

6.8 PIVUS

PIVUS is a community-based study where all men and women at age 70 living in Uppsala, Sweden were invited to participate in 2001 [90]. The 1,016 participants (50% women) have been extensively phenotyped including measurements of endothelial function and arterial compliance, cardiac function and structure by ultrasound and magnetic resonance imaging, evaluation of atherosclerosis by ultrasound and magnetic resonance imaging, seven day food intake recordings, detailed electrocardiogram (ECG) analysis, cardiovascular autonomic function and body composition by dual-energy X-ray absorptiometry.

Blood samples were drawn between 8 and 10 AM after an overnight fast. Plasma taken in EDTA-tubes was centrifuged, aliquoted and frozen within one hour. Serum was aliquoted and frozen within two hours. These samples were stored at -70° C in liquid nitrogen. Metabolomic profiling was performed on serum samples. The PIVUS study was used in study III. In total, 767 individuals were included in the analyses.

6.9 UK Biobank

UK Biobank recruitment took place between March 2007 and July 2010 via the UK National Health Service at 21 centres across England, Wales and Scotland [77]. All individuals aged 40-69 living within a reasonable travelling distance of an assessment

centre were asked to participate via postal invitation with a telephone follow-up. The overall response rate was 5.47 %.

Participants completed a "whole-body" assessment of 90-min duration that included a computerized questionnaire on lifestyle and medical history as well as blood and urine collection, a verbal interview with a trained nurse and physical measures including blood pressure, arterial stiffness, eye measures, hand grip strengths, anthropometry, bone densitometry of heel, spirometry and ECG from exercise test. The UK Biobank study was used in study IV. In total, 498,103 participants were included in the analyses.

7 Results and discussion

7.1 Prediction measures and sampling design

The high cost of laboratory assays makes the ascertainment of new omics technologies in all participants of large epidemiological studies unfeasible. It is therefore important to use sampling designs to determine which samples to select.

Study I investigates a particular aspect of this problem, that is, if prediction measures such as calibration, discrimination and reclassification can be calculated within traditional sampling designs and which sampling design is more appropriate to calculate such measures.

Specifically, we considered two sampling designs, case cohort and nested-case control (described in section 4.3), and two sampling schemes for each design:

1. Unstratified designs. For the case-cohort design, the sub-cohort is a random sample from the original cohort; for the nested case-control design, x controls are selected at random from individuals at risk at each case's failure time.
2. Stratified designs. For the case-cohort design, four strata (male or female and age higher or lower than the median) are considered and a number of participants proportional to the number of cases in each of these strata is randomly sampled from the original cohort. For each case in the nested case-control design, x controls at risk with the same sex and age (fine matching) are selected.

Results

We first compared the four combinations of sampling designs and sampling schemes for association with CHD in TwinGene. The gold standard was the association observed in the entire cohort. We found that both case-cohort and nested case-control designs gave more accurate results when stratified/matched sampling was used. Stratified case-cohort and matched nested case-control designs were comparable in terms of accuracy and efficiency of estimates; the unmatched nested case-control design was more efficient than the unstratified case-cohort design.

Secondly, we evaluated which sampling design more accurately calculated the risk of a CHD event within three years. To account for the sampling process, the individual risk was calculated by reweighting the baseline hazard with appropriate weights. Overall, the estimated individual risk from the sampling designs was comparable to

that observed in the entire cohort, except for the matched nested case-control design, which resulted in biased estimates.

Third, we compared sampling designs and whole cohort in terms of measures of discrimination (C-index), calibration (Grønnesby and Borgan goodness-of-fit test statistic) and reclassification (NRI). Overall, in the case-cohort designs and the unmatched nested case-control design, the estimates of the different prediction measures were similar to what was observed for the whole cohort, but the variability was higher. The C-index for the matched nested case-control design was lowest, underestimating the true value.

Discussion

Prediction measures can be calculated in a sampling design setting by using appropriate weights. Although this is straightforward for case-cohort designs where the subcohort is a random or stratified sample of the whole cohort, the appropriate weights for nested case-control designs are more complex as they are based on the inverse of the probability that a participant is ever selected as a control.

The finely matched nested-case control design obtained biased estimates of the individual risk and, consequently, the prediction measures were unsatisfactory. That is, failing to take into account the appropriate weights results in an underestimation of the C-index and consequently an overestimation of the discriminative power introduced by the additional marker.

More recently, it has been shown that is possible to obtain correct estimates of the individual risk from finely matched nested case control designs by using stratum-specific baseline hazards (Dr. Agus Salim, *personal communication*).

In conclusion, this study suggests that it is possible to calculate prediction measures from sampling designs and that the stratified case-cohort design is a reasonable choice both in terms of efficiency and simplicity in the calculation of prediction measures.

7.2 Genetics and prediction of cardiovascular disease

The utility of common genetic variants in primary prevention of cardiovascular disease has been subjected to several investigations [91, 92]. Overall, it has been suggested that there is not enough evidence for including genetics in current prediction models for cardiovascular disease [93]. Other studies have however

pointed out that useful levels of prediction may be approached with larger study samples [94].

In study II, we investigate the association between several MGRSs and CHD and evaluate their potential clinical utility. Two main MGRS were considered:

1. An overall MGRS obtained from 395 SNPs associated with CHD or CHD-related traits.
2. A CHD-specific MGRS obtained from 46 SNPs that have been found associated with CHD by the CARDIoGRAMplusC4D consortium, which is the largest genome-wide association studies meta-analysis on CHD to date [95].

Results

We first investigated the association between the MGRS and CHD after adjustment of established risk factors in 10,612 participants from six prospective studies experiencing 781 CHD events. Both the overall MGRS and the CHD-specific MGRS were highly significantly associated with CHD. Specifically, participants who were in the upper quartile of the distribution of the overall MGRS had 1.54 times [95% confidence interval (95% CI): 1.25-1.92] higher risk of CHD compared with individuals in the lowest quartile.

Second, we studied discrimination and reclassification. The overall MGRS significantly improved risk classification beyond established FHS risk factors (NRI= 4.2%), but the discrimination improvement was modest.

Third, we evaluated the number of events prevented and the number of event-free life years saved if the CHD-specific MGRS was measured in addition to the established risk factors. That is, assuming a risk reduction of 20% for individuals treated with statins [96], the targeted assessment of the genetic risk score among intermediate risk subjects could help to prevent about four additional CHD events during a ten-year period, which corresponds to one avoided event for every 318 people screened. If the CHD-specific genetic score was measured in the entire study population, 3.15 event-free life years per 1000 people screened would have been saved.

Discussion

The use of genetic profiling in primary prevention of CHD might improve allocation of patients to correct risk strata. A topical editorial [97] mentioning ours and other recent studies [98, 99] has highlighted how the prediction performances of genetic

risk scores are superior to those from other recently suggested biomarkers for prediction of cardiovascular disease (e.g. C-reactive protein, fibrinogen) [100] and wished for the translation of these findings in clinically meaningful action.

Genetic profiling has some advantages compared with other biomarkers. For example, genetic markers need to be measured only once and are likely to be predictive throughout life enabling earlier primary prevention in high-risk individuals. Moreover, the genotyping cost has rapidity dropped in the past years making genetic markers potentially cost-effective. However, formal studies investigating this aspect are needed. Finally, only carefully designed randomized trials will be able to provide a convincing confirmation about the expected benefits of genetics scores for primary prevention.

In conclusion, genetic information could be of some clinical value for prediction of CHD, although further studies are needed to address aspects such as feasibility, ethics, and cost efficiency of genetic profiling in the primary prevention setting.

7.3 Metabolomics and cardiovascular disease

The exploration of the metabolome holds a great potential to fuel the discovery of novel biomarkers of CHD. Such exploration has only become feasible in the past few years since technological advancements have allowed the measurements of hundreds of metabolites in thousands of samples. However, very few studies have performed a truly non-targeted metabolomics approach due to bioinformatics challenges and difficulties with the annotation of the metabolites.

In study III, we aimed to investigate the association between metabolic features and CHD, and to integrate genetic and metabolomics analysis to delineate the underlying biological mechanisms and evaluate potential causal effects of the novel biomarkers.

Results

We first studied the association between circulating metabolites and incident CHD events. We identified four metabolites that were associated with CHD after adjustment for established cardiovascular risk factors in 1,028 ULSAM participants (No. of events=131) and replicated in 1,670 TwinGene participants (No. of events=282). All four metabolites were lipid-related species: Two lysophosphatidylcholines, one sphingomyelin and one monoglyceride.

Second, we studied the joint discrimination and reclassification properties of the four metabolites in TwinGene. We observed a significantly improved risk classification beyond established FHS risk factors (NRI= 9.2%), but the discrimination improvement was modest.

Third, we explored the associations of the four novel metabolites with main cardiovascular risk factors, as well as with markers of oxidative stress, inflammation and subclinical CVD in additional 970 PIVUS participants. Lysophosphatidylcholines were negatively associated with BMI and with less evidence of subclinical CVD; a reverse pattern was observed for the monoglyceride.

Fourth, we studied the association with established CHD-SNPs and performed a Mendelian randomization approach. Only the monoglyceride showed an enrichment of significant associations with CHD-associated SNPs and a weak, but positive causal effect, as suggested by Mendelian randomization analysis.

Discussion

Four metabolites were identified as promising biomarkers of CHD: Two lysophosphatidylcholines, one sphingomyelin and one monoglyceride. While not much is known about the involvement of sphingomyelins in the pathogenesis of CHD, lysophosphatidylcholines has been previously indicated by functional studies to have a pro-inflammatory and pro-atherogenic effect [101]. This is in contrast with what observed in this study. However, recent population-based studies have suggested a protective effect of lysophosphatidylcholines on cardiovascular risk [75, 102], diabetes [103] and Alzheimer's disease [104]. We are currently performing zebrafish experiments to further delineate the role of these metabolites in lipid deposition and atherosclerotic plaques formation.

In conclusion, four lipid-related metabolites with evidence for clinical utility, as well as a causal role in CHD development were identified using a metabolomics approach. The use of animal models should help to further clarify the causal role of these metabolites.

7.4 Prediction of five-year overall mortality

It is an important public health priority and a central issue in clinical decision making to identify potential predictors of mortality, especially in the working age population.

Such predictors can be combined in a prediction score, which allows identification and risk stratification of individuals with reduced life expectancy.

Large, contemporary prospective studies collecting a wide range of measurements are needed to reach these goals. In study IV of this thesis, we compared more than 600 measurements of demographics, health and lifestyle with all-cause mortality and five cause-specific mortality categories. Moreover, we developed and validated a prediction score for five-year mortality using only self-reported information.

Results

We first studied the sex-specific association between 655 measurements and all-cause and cause-specific mortality in 498,103 UK Biobank participants experiencing 8,532 deaths during a median follow-up of five years. We found that measures that can simply be obtained by verbal interview without physical examination were the strongest predictors of all-cause mortality. Self-reported health and walking pace were among the strongest predictors in both genders and across different causes of deaths.

Second, we considered previously healthy individuals, by excluding those with a Charlson index greater than one. In this subgroup of individuals, smoking habits were the strongest category of mortality predictors.

Third, we built, validated and calibrated a prediction score. We selected 13 and 11 self-reported predictors for men and women, respectively, using a backward stepwise variable selection approach. Thus, we created a prediction model and validated it in the participants from the two Scottish centers, which were not used to develop the prediction score. The discrimination abilities were good (C-index 0.80 [95% CI 0.77-0.83] for men and 0.79 [95% CI 0.76-0.83] for women). Finally, we calibrated our prediction score using UK life tables and census information so that it was representative of the entire UK population aged 40 to 70.

Discussion

The study provides an extensive analysis of predictors of mortality in a large, contemporary prospective cohort study. By ranking measurements in terms of their predictive abilities we could provide new insights about the relative importance, as well as reveal some unexpected associations.

The most predictive self-reported measurements were combined in a prediction score that could be used by laypersons to improve self-awareness of the health status, by clinicians to identify high patients to target with specific interventions and by governmental and health organizations to decrease the burden of certain risk factors. To calibrate the prediction model, an innovative strategy that used life tables and census information was used. This allowed to overcome some of the issues related to the generalizability of the study.

Finally, all results were made available on an interactive website (www.ubble.co.uk) where it is possible to explore the observed associations in detail to generate new research hypotheses, and to calculate the biological age through an online questionnaire.

8 Future directions

8.1 Sampling designs and big data

With the advent of the 'big-data' era, one would expect to observe a growing interest in sampling designs. Instead, while much effort has been spent in developing strategies to optimize and parallelize computations, the research in sampling designs has remained limited to specialists in the field and failed to reach out to a larger audience.

Abundance of exceedingly technical literature, lack of comprehensive reviews on the topic and failure to implement analytical methods in statistical software are all factors contributing to the general lack of translation. For example, in R, the only function available to obtain estimators for case-cohort designs is the *cch* function from the *survival* package. This function has several limitations, including the inability to handle time-dependent covariates. More recently, the *survey* package from Thomas Lumely has implemented more complex sampling designs, but the functions are not optimized for big datasets [105].

Furthermore, most of the current research focuses on the correct estimation of standard errors, while much less attention has been dedicated to measures of predictive performances of the model, which are central in the 'big data' paradigm.

Indeed, it seems underappreciated the fact that a well-planned sampling design can deliver virtually identical information compared to the analysis conducted on the entire population, with the advantage of greatly reducing the computational burden. For example, sampling designs are routinely used in large population-based registries as an effective way to investigate epidemiological questions without the need to process the entire population [106-108].

Software implementation, translational research and better communication with researchers outside the narrow research field are ways to move forward research and applications of sampling designs.

8.2 Prediction models: Thinking outside the box

Newly discovered markers are often evaluated in addition to established risk factors for their ability to improve current prediction models. Although this is certainly a valid approach, it comes with the assumption that current risk factors are optimal

predictors. One can argue that this is not the case given that current prediction models have been developed in sample sizes that were smaller than those currently available and included a limited number of potential predictors.

With the advent of several new large initiatives in different countries, collecting baseline information from hundreds of thousands of individuals [77, 78], there is now room to rethink current prediction models. Novel untargeted screening of large risk factors collections can be used to identify new predictors. Such screening should include also those risk factors that are not causatively related with the outcome of interest and should identify uncorrelated risk factors with the aim to improve prediction.

Variable selection approaches using penalized likelihood can be use to achieve these goals and they have been shown to be more appropriate than traditional stepwise selection approaches [109, 110].

Moreover, with the widespread use of informatics technologies among patients and in the clinic, it is possible to start to think about non-traditional modelling strategies accounting for complex interactions and non-linear effects that can be directly integrated in online calculators and without the need to relay on printed score charts. For example, ensemble algorithms derived from the machine-learning field, has been shown to perform better than traditional linear models alone in epidemiological studies [111].

Finally, prediction models for the same outcome are often developed on different populations, hampering the comparison of their predictive performances. Open risk prediction competitions, such as the DREAM Challenges [112], can be used to obtain an objective comparison of different prediction models in the same population, to increase the generalizability and the transparency of the results and to motivate the researchers to think ‘outside the box’ in terms of statistical modelling.

Evidence-based medicine will see an increase use of prediction models as a way to deliver personalized and cost-effective care. There is a need to rethink current prediction models in the light of available methodologies and new large-scale efforts. Prediction competitions and faster implementation of novel prediction models in clinical guidelines are ways to motivate researchers to take on the challenge of delivering high-quality and innovative research in this field.

8.3 Clinical utility assessment: Understanding the context

The original NRI paper [31], published in 2008, has been cited more than 2,200 times and has generated a plethora of comments and critics [33, 113-115]. Researchers have been using this measure to claim the clinical utility of new markers for cardiovascular disease [116], breast cancer [117] and diabetes [118]. However, several misuses of this measure have been reported [119], especially in investigations of diseases where no clinically relevant risk categories are available. These critics have sorted the paradoxical effect of having authors explicitly mentioning the lack of NRI use in their paper [120, 121].

Flaws in claims of clinical utility are not new. In a lucid paper [122], Tzoulaki and colleagues have examined 79 studies evaluating improvement in predictive performance when predictors where added to the Framingham risk score. They found that although the majority of examined studies claimed additional predictive value beyond what the Framingham risk score could achieve, most had flaws in their design, analyses, and reporting. Several issues where related to the incorrect use of the AUC measure.

Taken together, these observations highlight the need for a better definition of which measures should be considered before claiming clinical utility. Leading experts in the field have published guidelines on how to evaluate predictive performance of new markers [123, 124], but it seems that most of these advises are not translated in practice. But most importantly, even in well-conducted studies, it is not clear what should be the natural step after reporting evidence of improved predictive performances.

It is therefore reasonable to argue that current measures of calibration, discrimination and reclassification are not directly translatable into clinical actions. Instead, these measures should mostly serve as screening tools to determine which markers should be taken forward to formal health economic evaluations and, when possible, to clinical trials.

Thus, it is important to conceptualize these measures within a broader directional framework that goes from marker discovery to introduction in clinical practice. They occupy a central position, following marker discovery but before appropriate investigations of clinical utility. Lacking of understanding this concept might result in wrong claims.

8.4 Integration of omics technologies using a vertical approach

Genome-wide association studies have paved the way in terms of statistical rigour, reproducibility, agnostic believes and data-sharing. Similar approaches have been used to interrogate other biological layers, such as methylation and metabolic profiles, and have been proved to be successful in discovering novel associations [103, 125].

These horizontal approaches are relatively easy to implement and have the advantages to provide new discoveries with a high grade of reproducibility.

However, the translation of such findings into mechanistic insights, clinically relevant results and drug discovery is more likely to be achieved by integrating different omics layers including in vivo and in vitro follow-up.

Such vertical approach, as opposed to the horizontal approach of single omics screening, should be able to prioritize pathways of interest by both data-driven discovery and integration of external information.

While horizontal screenings are hypothesis-free and relies only on information derived from the data, an integrative vertical approach needs to take into account external information regarding the biological nature of the interaction between omics layers.

This information can be partially obtained by integrating data provided by large consortia, such as GTEx [126], ENCODE [127] and the Human Protein Atlas [128].

Moreover, information from previous horizontal screenings (e.g. summary statistics from GWAS meta-analyses) should also be included in order to prioritize specific pathways of interest and reduce the search space.

The complexity of the vertical approaches is far higher than that of horizontal approaches and several improvements need to be made.

First, statistical methods able to integrate different layers of omics data and to handle constrains of biological nature are needed. A recently proposed method goes in this direction by getting inspiration from the field of geostatistics [129]. However, no external information can be yet integrated in this model.

Second, deeply phenotyped studies need to be collected. Repeated measurements should also be assessed in order to estimate the inherent variability due to environment influence and physiological states. Moreover, new ways to collect tissues other than blood and urine should be explored.

Finally, high-throughput in vivo and in vitro experiments need to be conducted in a fast and cost-effective manner to validate findings and allow for direct therapeutic translation.

In conclusion, horizontal and vertical approaches should be combined to discover new clinically and therapeutically relevant targets. The latter approach poses several challenges that need to be addressed by future research.

8.5 Data sharing and privacy: Opportunities and challenges

Data sharing is of enormous importance in modern science. First, it allows the access by a larger scientific audience to resources that would have otherwise been available to few researchers. Second, it increases the pace of scientific discoveries by allowing data to be used for purposes other than what they have been collected. Third, it allows the scientific community to scrutinize the results from a study, thus increasing the validity of its findings.

For example, using publically available microarray experiments researchers have recently identified a promising therapeutic target for diabetes [130]. Data sharing have also stimulated a wealth of statistical methods specifically designed to handle aggregate data, such as summary statistics from GWAS [131].

There are numerous sources of publically available data. Omics data repositories have been created in parallel with the increased demand for data sharing from scientific journals, as a way to improve reproducibility and transparency. Data from clinical trials, independently if they have been successful or not, are increasingly being made publically available allowing the possibility to use this data for new and unexpected research questions [132]. Large initiatives such as the Global Alliance for Genomics and Health have been created to make possible the large-scale collection of data on genome sequencing and clinical outcomes within a common framework and using standardized methodology. Finally, for profit organizations such as 23andMe have recognized the importance of sharing their data with the scientific community as a way to stimulate innovative ideas and identify new commercially viable targets.

On the other side, with the increased availability of medical data on the web, privacy concerns have been raised. For example, anonymized genomics information can be combined with publically available genealogy databases to obtain personal information (e.g. patient's surname) without patient's consent [133]. Similarly, it has

been shown that DNA of individuals can be identified from complex mixtures of information, for example from GWAS summary statistics [134].

It is therefore important to engage the scientific community and the general public in a rational discussion about the risks and benefits of having medical and genetics information publically available. As a way to achieve this goal, scientists should start by strengthening their ability to communicate scientific findings to laypersons, for example by employing tools that make results interpretable by a broader community (e.g. interactive websites, mobile applications).

In conclusion, data sharing is beneficial to the scientific community. Improved communication and a clear, balanced informed consent are key elements to uphold this benefit together with a long-lasting credibility in the general public.

10. Acknowledgments

This thesis and most of what I accomplished in these years has been possible thanks to the impressionistic contribution of many people, each in a distinctive fashion.

First, a special mention goes to my PhD advisor Erik Ingelsson. You have been a role model for me and stand for all the qualities a supervisor should have. Every comment from you is constructive and enriching, which is a rare quality. Thank you for keeping me sharp and my vision open.

Also, my co-advisors, Patrik Magnusson and Marie Reilly, it has been an honour to work with you. Patrik, your broad knowledge and critical spirit are unique. Marie, your positive attitude, kindness and sampling-design knowledge have have made my work worth the effort in the past four years.

Moreover, I want to thank Yudi Pawitan. I have learned so much by working with you that I'm not afraid calling myself a statistician. And I thank Rino Bellocco for giving me the opportunity to write my master thesis at MEB and firstly introducing me to Erik.

Marcel Den Hoed, Tove Fall, Samira Salihovic, Sara Hägg, Ci Song, Stefan Gustaffson, Jitender Kumar, Åsa Hedman and all the new and old members of the Ingelsson's group. Thank you for all your help and constructive discussions; this has been a terrific and fun collaboration!

Thanks to my collaborators in Rotterdam, Cecile Janssens and Henning Tiemeier. It has been a productive and inspiring period.

I want to thank all the peers in the *biostat* group. This is a fantastic and unique group. We have the privilege of discussing theories, where most people just apply them.

MEB has been a balanced and excellent place to work, this is also thanks to: Hatef Darabi, Edoardo Colzani, Paul Dickman, Andrea Foebel, Keith Humphreys, Caroline

Weibull, Therese Andersson, Flaminia Chiesa, Mark Clements, Daniela Mariosa, Alessandra Grotta, Anne Örtqvist, Rikard Öberg, Zack Yusof, Fei Yang, Robert Szulkin, Robert Karlsson, Sven Sandin, Mattias Rantalainen, Alexander Ploner, Arvid Sjölander, Tracy Peters, Nancy Pedersen, Jingmei Li, Donghwan Lee, Andreas Karlsson, Anna Johansson, Vilhelmina Ullemar, Alexandra Jauhiainen, Marie Jansson, Miriam Elfström and Elin Ross.

Thanks to Alessandro Furlan, Vilma Rraklli, Mustafa and Haythem Ismail Muhammad, Cecile Zaouter, Giulia Gaudenzi, Tina Heyder, Fausto Molinari, Karina Hentrich and all those who made the dark nights brighter and the Swedish winters warmer.

Special thanks to my family, who have been always supportive. Even if you couldn´t see me very often, you always helped me keeping my focus. Grazie!

Lastly, Steffi, I would like to thank you for your support in my career and in my life. For making this possible certain choices had to be made and I´m grateful and proud on your positive attitude and open mind.

11. References

[1] Dawber TR, Meadors GF, Moore FE, Jr. Epidemiological approaches to heart disease: the Framingham Study. *American journal of public health and the nation's health.* 1951;41(3):279-81.

[2] Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet.* 2012;380(9859):2095-128.

[3] Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS medicine.* 2006;3(11):e442.

[4] Ingelsson E. The grand challenge of cardiovascular epidemiology: Turning the tide. *Frontiers in Cardiovascular Medicine.* 2014;1.

[5] Murray CJ, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet.* 2012;380(9859):2197-223.

[6] Mendis S, Puska P, Norrving B, et al. *Global atlas on cardiovascular disease prevention and control*. Geneva: World Health Organization in collaboration with the World Heart Federation and the World Stroke Organization; 2011.

[7] Kannel WB, Dawber TR, McGee DL. Perspectives on systolic hypertension. The Framingham study. *Circulation.* 1980;61(6):1179-82.

[8] Thomas HE, Jr., Kannel WB, Dawber TR, et al. Cholesterol-phospholipid ratio in the prediction of coronary heart disease. The Framingham study. *The New England journal of medicine.* 1966;274(13):701-5.

[9] Kannel WB, McGee DL. Diabetes and cardiovascular disease. The Framingham study. *Jama.* 1979;241(19):2035-8.

[10] Kannel WB, LeBauer EJ, Dawber TR, et al. Relation of body weight to development of coronary heart disease. The Framingham study. *Circulation.* 1967;35(4):734-44.

[11] Friedman LA, Kimball AW. Coronary heart disease mortality and alcohol consumption in Framingham. *American journal of epidemiology.* 1986;124(3):481-9.

[12] Doyle JT, Dawber TR, Kannel WB, et al. Cigarette smoking and coronary heart disease. Combined experience of the Albany and Framingham studies. *The New England journal of medicine.* 1962;266:796-801.

[13] Kannel WB. Habitual level of physical activity and risk of coronary heart disease: the Framingham study. *Canadian Medical Association journal.* 1967;96(12):811-2.

[14] Lloyd-Jones DM, Nam BH, D'Agostino RB, Sr., et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *Jama.* 2004;291(18):2204-11.

[15] Zdravkovic S, Wienke A, Pedersen NL, et al. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *Journal of internal medicine.* 2002;252(3):247-54.

[16] Kannel WB. Some lessons in cardiovascular epidemiology from Framingham. *The American journal of cardiology.* 1976;37(2):269-82.

[17] Levi F, Lucchini F, Negri E, et al. Trends in mortality from cardiovascular and cerebrovascular diseases in Europe and other areas of the world. *Heart.* 2002;88(2):119-24.
[18] Ford ES, Greenlund KJ, Hong Y. Ideal cardiovascular health and mortality from all causes and diseases of the circulatory system among adults in the United States. *Circulation.* 2012;125(8):987-95.
[19] Yang Q, Cogswell ME, Flanders WD, et al. Trends in cardiovascular health metrics and associations with all-cause and CVD mortality among US adults. *Jama.* 2012;307(12):1273-83.
[20] Biesheuvel CJ, Vergouwe Y, Oudega R, et al. Advantages of the nested case-control design in diagnostic research. *BMC medical research methodology.* 2008;8:48.
[21] Pepe MS. *The statistical evaluation of medical tests for classification and prediction*. Oxford ; New York: Oxford University Press; 2003.
[22] Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Statistics in medicine.* 2006;25(20):3474-86.
[23] Uno H, Cai T, Pencina MJ, et al. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in medicine.* 2011;30(10):1105-17.
[24] Harrell FE, Jr., Califf RM, Pryor DB, et al. Evaluating the yield of medical tests. *Jama.* 1982;247(18):2543-6.
[25] Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation.* 2007;115(7):928-35.
[26] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44(3):837-45.
[27] Vickers AJ, Cronin AM, Begg CB. One statistical test is sufficient for assessing new predictive markers. *BMC medical research methodology.* 2011;11:13.
[28] Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd ed. New York: Wiley; 2000.
[29] May S, Hosmer DW. A simplified method of calculating an overall goodness-of-fit test for the Cox proportional hazards model. *Lifetime data analysis.* 1998;4(2):109-20.
[30] Steyerberg EW. *Clinical prediction models : a practical approach to development, validation, and updating*. New York, NY: Springer; 2009.
[31] Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in medicine.* 2008;27(2):157-72; discussion 207-12.
[32] Mihaescu R, van Zitteren M, van Hoek M, et al. Improvement of risk prediction by genomic profiling: reclassification measures versus the area under the receiver operating characteristic curve. *American journal of epidemiology.* 2010;172(3):353-61.
[33] Greenland S. The need for reorientation toward cost-effective prediction: comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). *Statistics in medicine.* 2008;27(2):199-206.

[34] Pepe MS, Janes H. Commentary: Reporting standards are needed for evaluations of risk reclassification. *International journal of epidemiology.* 2011;40(4):1106-8.
[35] Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *The New England journal of medicine.* 1980;302(20):1109-17.
[36] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2006;26(6):565-74.
[37] Rapsomaniki E, White IR, Wood AM, et al. A framework for quantifying net benefits of alternative prognostic models. *Statistics in medicine.* 2012;31(2):114-30.
[38] Kattan MW, Cowen ME. *Encyclopedia of medical decision making*. Thousand Oaks, Calif.: SAGE Publications; 2009.
[39] Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham Study. *The American journal of cardiology.* 1976;38(1):46-51.
[40] Anderson KM, Odell PM, Wilson PW, et al. Cardiovascular disease risk profiles. *American heart journal.* 1991;121(1 Pt 2):293-8.
[41] D'Agostino RB, Sr., Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation.* 2008;117(6):743-53.
[42] Schnabel RB, Sullivan LM, Levy D, et al. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *Lancet.* 2009;373(9665):739-45.
[43] Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). *Jama.* 2001;285(19):2486-97.
[44] Pencina MJ, D'Agostino RB, Sr., Larson MG, et al. Predicting the 30-year risk of cardiovascular disease: the framingham heart study. *Circulation.* 2009;119(24):3078-84.
[45] Wilson PW, Meigs JB, Sullivan L, et al. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Archives of internal medicine.* 2007;167(10):1068-74.
[46] Parikh NI, Pencina MJ, Wang TJ, et al. A risk score for predicting near-term incidence of hypertension: the Framingham Heart Study. *Annals of internal medicine.* 2008;148(2):102-10.
[47] D'Agostino RB, Wolf PA, Belanger AJ, et al. Stroke risk profile: adjustment for antihypertensive medication. The Framingham Study. *Stroke; a journal of cerebral circulation.* 1994;25(1):40-3.
[48] Perk J, De Backer G, Gohlke H, et al. European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). *European heart journal.* 2012;33(13):1635-701.
[49] Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *Bmj.* 2008;336(7659):1475-82.

[50] Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study. *Circulation.* 2002;105(3):310-5.
[51] Voss R, Cullen P, Schulte H, et al. Prediction of risk of coronary events in middle-aged men in the Prospective Cardiovascular Munster Study (PROCAM) using neural networks. *International journal of epidemiology.* 2002;31(6):1253-62; discussion 62-64.
[52] Woodward M, Brindle P, Tunstall-Pedoe H, et al. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart.* 2007;93(2):172-6.
[53] Ferrario M, Chiodini P, Chambless LE, et al. Prediction of coronary events in a low incidence population. Assessing accuracy of the CUORE Cohort Study prediction equation. *International journal of epidemiology.* 2005;34(2):413-21.
[54] Wu Y, Liu X, Li X, et al. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. *Circulation.* 2006;114(21):2217-25.
[55] Saito I, Sato S, Nakamura M, et al. A low level of C-reactive protein in Japanese adults and its association with cardiovascular risk factors: the Japan NCVC-Collaborative Inflammation Cohort (JNIC) study. *Atherosclerosis.* 2007;194(1):238-44.
[56] Sritara P, Cheepudomwit S, Chapman N, et al. Twelve-year changes in vascular risk factors and their associations with mortality in a cohort of 3499 Thais: the Electricity Generating Authority of Thailand Study. *International journal of epidemiology.* 2003;32(3):461-8.
[57] Knuiman MW, Vu HT, Bartholomew HC. Multivariate risk estimation for coronary heart disease: the Busselton Health Study. *Australian and New Zealand journal of public health.* 1998;22(7):747-53.
[58] U.S. Preventive Services Task Force. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of internal medicine.* 2008;149(9):627-37.
[59] Lee SJ, Boscardin WJ, Stijacic-Cenzer I, et al. Time lag to benefit after screening for breast and colorectal cancer: meta-analysis of survival data from the United States, Sweden, United Kingdom, and Denmark. *Bmj.* 2013;346:e8441.
[60] Lee SJ, Leipzig RM, Walter LC. Incorporating lag time to benefit into prevention decisions for older adults. *Jama.* 2013;310(24):2609-10.
[61] Yourman LC, Lee SJ, Schonberg MA, et al. Prognostic indices for older adults: a systematic review. *Jama.* 2012;307(2):182-92.
[62] Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases.* 1987;40(5):373-83.
[63] Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409(6822):860-921.
[64] Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science.* 2001;291(5507):1304-51.
[65] International HapMap Consortium. A haplotype map of the human genome. *Nature.* 2005;437(7063):1299-320.

[66] Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends in genetics : TIG.* 2001;17(9):502-10.
[67] Lander ES. Initial impact of the sequencing of the human genome. *Nature.* 2011;470(7333):187-97.
[68] Park JH, Wacholder S, Gail MH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature genetics.* 2010;42(7):570-5.
[69] Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Human molecular genetics.* 2009;18(18):3525-31.
[70] Janssens AC, Aulchenko YS, Elefante S, et al. Predictive testing for complex diseases using multiple genes: fact or fiction? *Genetics in medicine : official journal of the American College of Medical Genetics.* 2006;8(7):395-400.
[71] Wray NR, Yang J, Hayes BJ, et al. Pitfalls of predicting complex traits from SNPs. *Nature reviews Genetics.* 2013;14(7):507-15.
[72] Nicholson JK, Holmes E, Kinross JM, et al. Metabolic phenotyping in clinical and surgical environments. *Nature.* 2012;491(7424):384-92.
[73] Tzoulaki I, Ebbels TM, Valdes A, et al. Design and analysis of metabolomics studies in epidemiologic research: a primer on -omic technologies. *American journal of epidemiology.* 2014;180(2):129-39.
[74] Shah SH, Sun JL, Stevens RD, et al. Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease. *American heart journal.* 2012;163(5):844-50 e1.
[75] Stegemann C, Pechlaner R, Willeit P, et al. Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study. *Circulation.* 2014;129(18):1821-31.
[76] Vaarhorst AA, Verhoeven A, Weller CM, et al. A metabolomic profile is associated with the risk of incident coronary heart disease. *American heart journal.* 2014;168(1):45-52 e7.
[77] Palmer LJ. UK Biobank: bank on it. *Lancet.* 2007;369(9578):1980-2.
[78] Stolk RP, Rosmalen JG, Postma DS, et al. Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. *European journal of epidemiology.* 2008;23(1):67-74.
[79] Chao C, Song Y, Cook N, et al. The lack of utility of circulating biomarkers of inflammation and endothelial dysfunction for type 2 diabetes risk prediction among postmenopausal women: the Women's Health Initiative Observational Study. *Archives of internal medicine.* 2010;170(17):1557-65.
[80] Kim HC, Greenland P, Rossouw JE, et al. Multimarker prediction of coronary heart disease risk: the Women's Health Initiative. *Journal of the American College of Cardiology.* 2010;55(19):2080-91.
[81] Sattar N, Welsh P, Sarwar N, et al. NT-proBNP is associated with coronary heart disease risk in healthy older women but fails to enhance prediction beyond established risk factors: results from the British Women's Heart and Health Study. *Atherosclerosis.* 2010;209(1):295-9.
[82] Kulathinal S, Karvanen J, Saarela O, et al. Case-cohort design in practice - experiences from the MORGAM Project. *Epidemiologic perspectives & innovations : EP+I.* 2007;4:15.
[83] Magnusson PK, Almqvist C, Rahman I, et al. The Swedish Twin Registry: establishment of a biobank and other recent developments. *Twin Res Hum Genet.* 2013;16(1):317-29.

[84] Anckarsater H, Lundstrom S, Kollberg L, et al. The Child and Adolescent Twin Study in Sweden (CATSS). *Twin research and human genetics : the official journal of the International Society for Twin Studies.* 2011;14(6):495-508.
[85] Pedersen NL, Friberg L, Floderus-Myrhed B, et al. Swedish early separated twins: identification and characterization. *Acta geneticae medicae et gemellologiae.* 1984;33(2):243-50.
[86] Finkel D, Reynolds CA, McArdle JJ, et al. Genetic variance in processing speed drives variation in aging of spatial and memory abilities. *Developmental psychology.* 2009;45(3):820-34.
[87] McClearn GE, Johansson B, Berg S, et al. Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science.* 1997;276(5318):1560-3.
[88] Gold CH, Malmberg B, McClearn GE, et al. Gender and health: a study of older unlike-sex twins. *The journals of gerontology Series B, Psychological sciences and social sciences.* 2002;57(3):S168-76.
[89] Byberg L, Siegbahn A, Berglund L, et al. Plasminogen activator inhibitor-1 activity is independently related to both insulin sensitivity and serum triglycerides in 70-year-old men. *Arteriosclerosis, thrombosis, and vascular biology.* 1998;18(2):258-64.
[90] Lind L, Fors N, Hall J, et al. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. *Arteriosclerosis, thrombosis, and vascular biology.* 2005;25(11):2368-75.
[91] Ripatti S, Tikkanen E, Orho-Melander M, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet.* 2010;376(9750):1393-400.
[92] Paynter NP, Chasman DI, Pare G, et al. Association between a literature-based genetic risk score and cardiovascular events in women. *Jama.* 2010;303(7):631-7.
[93] Redberg RF, Benjamin EJ, Bittner V, et al. ACCF/AHA 2009 performance measures for primary prevention of cardiovascular disease in adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures (Writing Committee to Develop Performance Measures for Primary Prevention of Cardiovascular Disease) developed in collaboration with the American Academy of Family Physicians; American Association of Cardiovascular and Pulmonary Rehabilitation; and Preventive Cardiovascular Nurses Association: endorsed by the American College of Preventive Medicine, American College of Sports Medicine, and Society for Women's Health Research. *Journal of the American College of Cardiology.* 2009;54(14):1364-405.
[94] Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS genetics.* 2013;9(3):e1003348.
[95] CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature genetics.* 2013;45(1):25-33.
[96] Cholesterol Treatment Trialists Collaborators, Mihaylova B, Emberson J, et al. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *Lancet.* 2012;380(9841):581-90.

[97] Thanassoulis G, Peloso GM, O'Donnell CJ. Genomic medicine for improved prediction and primordial prevention of cardiovascular disease. *Arteriosclerosis, thrombosis, and vascular biology.* 2013;33(9):2049-50.
[98] Isaacs A, Willems SM, Bos D, et al. Risk scores of common genetic variants for lipid levels influence atherosclerosis and incident coronary heart disease. *Arteriosclerosis, thrombosis, and vascular biology.* 2013;33(9):2233-9.
[99] Tikkanen E, Havulinna AS, Palotie A, et al. Genetic risk prediction and a 2-stage risk screening strategy for coronary heart disease. *Arteriosclerosis, thrombosis, and vascular biology.* 2013;33(9):2261-6.
[100] Emerging Risk Factors Collaboration, Kaptoge S, Di Angelantonio E, et al. C-reactive protein, fibrinogen, and cardiovascular disease prediction. *The New England journal of medicine.* 2012;367(14):1310-20.
[101] Schmitz G, Ruebsaamen K. Metabolism and atherogenic disease association of lysophosphatidylcholine. *Atherosclerosis.* 2010;208(1):10-8.
[102] Fernandez C, Sandin M, Sampaio JL, et al. Plasma lipid composition and risk of developing cardiovascular disease. *PloS one.* 2013;8(8):e71846.
[103] Wang-Sattler R, Yu Z, Herder C, et al. Novel biomarkers for pre-diabetes identified by metabolomics. *Molecular systems biology.* 2012;8:615.
[104] Karayalcin B, Aras G, Erbay G, et al. Parasternal lymphoscintigraphy using 99Tcm-dextran. *Nuclear medicine communications.* 1988;9(9):657-62.
[105] Lumley T. *Complex surveys : a guide to analysis using R*. Hoboken, N.J.: John Wiley; 2010.
[106] Holmes MD, Olsson H, Pawitan Y, et al. Aspirin intake and breast cancer survival - a nation-wide study using prospectively recorded data in Sweden. *BMC cancer.* 2014;14:391.
[107] Lagiou P, Hsieh CC, Trichopoulos D, et al. Neonatal growth and breast cancer risk in adulthood. *British journal of cancer.* 2008;99(9):1544-8.
[108] Fang F, Ye W, Fall K, et al. Loss of a child and the risk of amyotrophic lateral sclerosis. *American journal of epidemiology.* 2008;167(2):203-10.
[109] Harrell FE. *Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis*. New York: Springer; 2001.
[110] Tibshirani R. Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B Met.* 1996;58(1):267-88.
[111] Rose S. Mortality risk score prediction in an elderly population using machine learning. *American journal of epidemiology.* 2013;177(5):443-52.
[112] Marbach D, Costello JC, Kuffner R, et al. Wisdom of crowds for robust gene network inference. *Nature methods.* 2012;9(8):796-804.
[113] Cook NR. Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). *Statistics in medicine.* 2008;27(2):191-5.
[114] Van Calster B, Van Huffel S. Integrated discrimination improvement and probability-sensitive AUC variants. *Statistics in medicine.* 2010;29(2):318-9.
[115] Ware JH, Cai T. Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). *Statistics in medicine.* 2008;27(2):185-7.
[116] Ingelsson E, Schaefer EJ, Contois JH, et al. Clinical utility of different lipid measures for prediction of coronary heart disease in men and women. *Jama.* 2007;298(7):776-85.

[117] Rodenhuis S, Mandjes IA, Wesseling J, et al. A simple system for grading the response of breast cancer to neoadjuvant chemotherapy. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO.* 2010;21(3):481-7.
[118] Salomaa V, Havulinna A, Saarela O, et al. Thirty-one novel biomarkers as predictors for clinically incident diabetes. *PloS one.* 2010;5(4):e10100.
[119] Kerr KF, Wang Z, Janes H, et al. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology.* 2014;25(1):114-21.
[120] Goldstein BA, Knowles JW, Salfati E, et al. Simple, standardized incorporation of genetic risk into non-genetic risk prediction tools for complex traits: coronary heart disease as an example. *Frontiers in genetics.* 2014;5:254.
[121] Nielsen JB, Graff C, Rasmussen PV, et al. Risk prediction of cardiovascular death based on the QTc interval: evaluating age and gender differences in a large primary care population. *European heart journal.* 2014;35(20):1335-44.
[122] Tzoulaki I, Liberopoulos G, Ioannidis JP. Assessment of claims of improved prediction beyond the Framingham risk score. *Jama.* 2009;302(21):2345-52.
[123] Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart.* 2012;98(9):683-90.
[124] Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010;21(1):128-38.
[125] Dick KJ, Nelson CP, Tsaprouni L, et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet.* 2014;383(9933):1990-8.
[126] GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nature genetics.* 2013;45(6):580-5.
[127] ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57-74.
[128] Uhlen M, Oksvold P, Fagerberg L, et al. Towards a knowledge-based Human Protein Atlas. *Nature biotechnology.* 2010;28(12):1248-50.
[129] Wheeler HE, Aquino-Michaels K, Gamazon ER, et al. Poly-omic prediction of complex traits: OmicKriging. *Genetic epidemiology.* 2014;38(5):402-15.
[130] Kodama K, Toda K, Morinaga S, et al. Anti-CD44 Antibody Treatment Lowers Hyperglycemia and Improves Insulin Resistance, Adipose Inflammation, and Hepatic Steatosis in Diet-Induced Obese Mice. *Diabetes.* 2014.
[131] Yang J, Lee SH, Goddard ME, et al. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics.* 2011;88(1):76-82.
[132] Eichler HG, Abadie E, Breckenridge A, et al. Open clinical trial data for all? A view from regulators. *PLoS medicine.* 2012;9(4):e1001202.
[133] Gymrek M, McGuire AL, Golan D, et al. Identifying personal genomes by surname inference. *Science.* 2013;339(6117):321-4.
[134] Homer N, Szelinger S, Redman M, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS genetics.* 2008;4(8):e1000167.